Object Detection for Random Bin Picking using Point Pair Features

Wim Abbeloos and Toon Goedemé
KU Leuven, Department of Electrical Engineering

Point pair features are a popular representation for free-form 3D object detection and pose estimation. We investigate their performance in an industrial random bin picking context. A new method to automatically generate representative synthetic datasets, with a high degree of clutter and the presence of self similar features, is proposed. A simple heuristic method is used to reduce the computational complexity and to improve the speed and robustness.

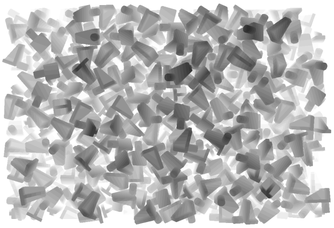

Fig. 1: Synthetically generated range image (darker colors are closer).

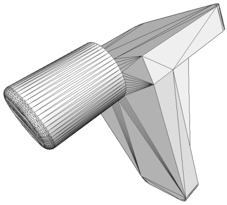

Fig. 2: Input 3D object model.

1 Synthetic Random Bin Picking Dataset

To obtain realistic and representative test data for random bin picking, synthetic scenes (Fig. 1) are automatically generated from 3D object models (Fig. 2). A physics simulation library is used to simulate dropping the object models into a bin. To reduce simulation times, the objects are decomposed into convex parts using Approximate Convex Decomposition. The final object poses are stored as the ground truth, and range and intensity images of the final scene are rendered.

2 Point Pair Feature based Object Detection

Point pair features [Drost et al., CVPR 2010] describe the relative position and orientation of points on the surface of an object. For two points $\mathbf{m_1}$ and $\mathbf{m_2}$ with normals $\mathbf{n_1}$ and $\mathbf{n_2}$, we set $\mathbf{d} = \mathbf{m_2} - \mathbf{m_1}$ and define the feature $\mathbf{F}$ as:

$$\mathbf{F}(\mathbf{m_1}, \mathbf{m_2}) = (\|\mathbf{d}\|_2, \angle(\mathbf{n_1}, \mathbf{d}), \angle(\mathbf{n_2}, \mathbf{d}), \angle(\mathbf{n_1}, \mathbf{n_2}))$$

with $\angle(\mathbf{a}, \mathbf{b}) \in [0 \quad \pi]$ the angle between two vectors (Fig. 3).
To learn a new object model, the feature vectors are discretized and stored in a hashtable. To detect an object in the scene, a uniform subsampling procedure is applied to the scene points. The point pair features between a set of randomly selected reference points and the subsampled points are calculated and compared to the ones in the hashtable. Matching vectors vote for an object pose. Similar poses are clustered and their poses are averaged.

3 Heuristic Hypothesis Generation

A heuristic approach is used to drastically reduce the complexity of detecting objects using PPFs, as it is not feasible to detect small objects in large pointclouds directly. A low pass filter is applied to the range image, and iteratively the closest point is selected (Fig. 4). The assumption made is that the highest objects will be the easiest to detect and grasp, as they are less likely to be occluded by other objects.

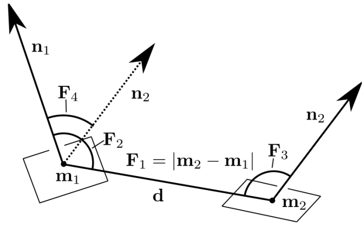

Fig. 3: Two surface points $\mathbf{m_i}$ and their normals $\mathbf{n_i}$ determine a point pair feature.

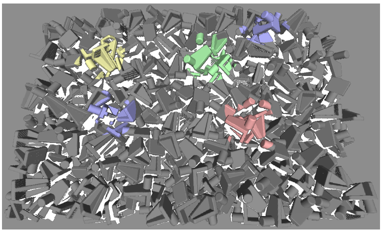

Fig. 4: The selected hypotheses, with the five highest points in color.

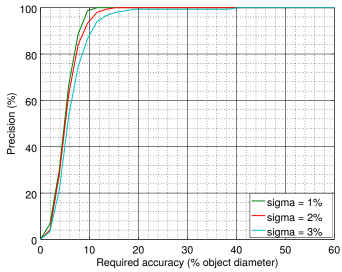

Fig. 5: Varying translation threshold.

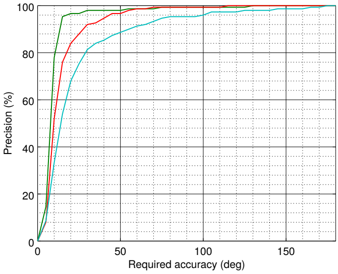

Fig. 6: Varying rotation threshold.

4 Experimental Results

A synthetic dataset of 30 bins, each containing 350 objects, was generated and the five highest points in each scene were selected as hypotheses. Different levels of noise were applied (as a percentage of the object diameter). The PPF based detection was used on each hypothesis and the resulting pose compared to the ground truth (without using Iterative Closest Point).

The PPFs were discretized into 30 steps (12° per step) for the normal angles and 20 steps for the distance, with the maximum distance corresponding to the object diameter. The subsampling resolution for the model and scene was set to five percent of the object diameter. All poses within a distance of 0.75 and a rotation angle of 20 degrees were clustered. The average detection time per hypothesis was 726 milliseconds using 20% of the subsampled scene points as reference points and only 70 milliseconds when using 2%.

The translation error (Fig. 5) is defined as the Euclidean distance between the detection and ground truth pose as a percentage of the object model diameter. The rotational error (Fig. 6) is the angle obtained by calculating the rotation matrix to align the frames of the detection and ground truth pose and converting it to Angle-Axis representation.